**Explore restaurants in Toronto**

Bharat Bhushan

Jan 19, 2021

**1. Introduction**

**1.1 Background**

Toronto is the capital city of the Canadian province of Ontario. it is the most populous city in Canada and the fourth most populous city in North America. Also, it is one of the main hub of business and tourism in Canada attracting lots of tourists from all around the world. As a part of Applied Capstone part in IBM Data Science course, I worked on this project where I explored different neighbourhoods in Toronto so as to get recommendations for a person planning to open a new restaurant in Toronto.

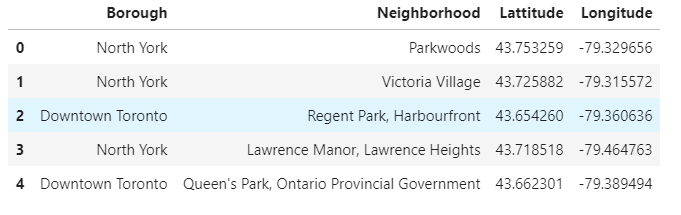
**1.2 Problem**

To explore different neighbourhoods in Toronto city using data fetched from Foursquare like list of restaurants in the neighbourhood, their categories and count in each neighbourhood and come up with useful insights to help a person who wants to open a new restaurant in Toronto in taking some of key decisions e.g. location, restaurant category.

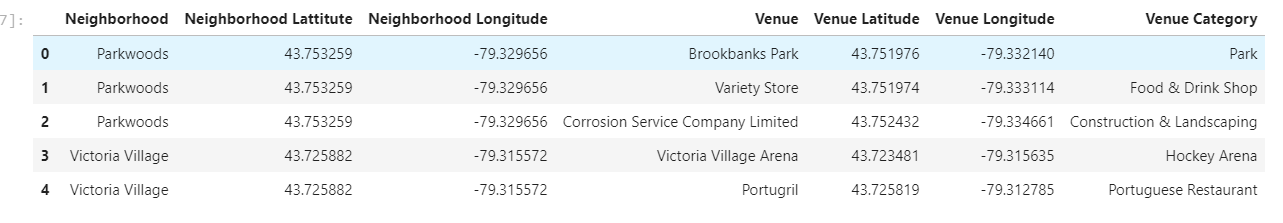
**2. Data collection and cleaning**

**2.1 Data sources**

I used a python web scraping code to collect information of different neighborhoods in Toronto from wiki page : [**Toronto Postal Codes**](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). Then, I used excel containing co-ordinates of each neighbourhood in Toronto provided in IBM Data Science Capstone course on Coursera and combined this data to create a dataframe :



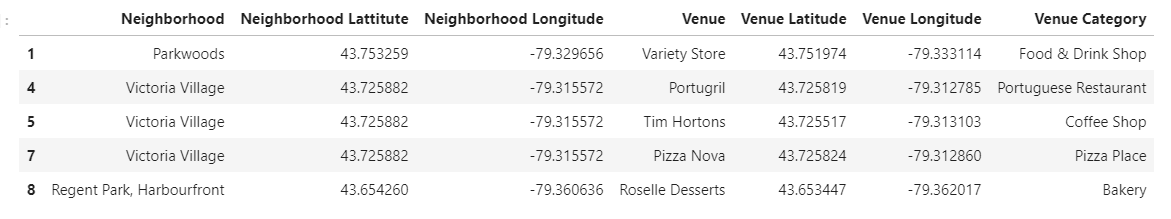
Using this data, I used Foursquare API to find venues of different categories in each neighbourhood in Toronto. It gave a dataset of 2100 venues :



**2.2 Data Cleaning**

As we are interested in restaurants business, next step is to segregate the data related to restaurants from all the venues.

Inspected the venue categories and found that most of the restaurants have a keyword from list of words e.g. ‘Diner’,’Coffee’,Pizza etc. Spent some time in identifying the keywords and then filtered data of all venues in Toronto to data related to restaurants. Final data has 1070 restaurants in Toronto city located in different neighbourhoods.



The data can be further analysed and transformed to derive the features needed to be fed to an unsupervised machine learning algorithm to generate groups of similar neighbourhoods.

a neighbourhood from each group are also discussed.